

1. A method for identifying a disease-influencing gene, the method comprising the steps of:
- a) selecting individuals having a risk factor for a disease;
  - b) creating queries regarding the individuals' behaviors and environments;
  - c) storing the queries on a server;
  - d) providing each of the individuals with a remotely programmable apparatus having a user interface for communicating the queries and for receiving responses, and having communication means for communicating with the server through a communication network;
  - e) transmitting the queries from the server to each of the remotely programmable apparatuses;
  - f) transmitting the responses of the individuals to the queries from the remotely programmable apparatuses to the server;
  - g) creating a database of the individuals' behaviors and environments;
  - h) using data mining techniques to distinguish a group of individuals having similar behavioral and environmental profiles;

- R and*
- i) categorizing the group of individuals into at least two categories according to the individuals' disease progression;
  - j) determining [the genotypes of] at least one physiologic characteristic relating to a genotype for each of the at least two categories of individuals;
  - k) using data mining techniques to find a gene difference between the at least two categories of individuals based upon the at least one physiologic characteristic relating to the genotype for each of the categories of individuals; and
  - l) identifying the disease-influencing gene.

- A2*
- 10. A method for identifying a disease-influencing gene, the method comprising the steps of:
    - a) selecting individuals having a risk factor for a disease;
    - b) creating queries regarding the individuals' behaviors and environments;
    - c) storing the queries on a server;
    - d) providing each of the individuals with a remotely programmable apparatus having a user interface for communicating the queries and for receiving responses,

and having communication means for communicating with the server through a communication network;

- e) transmitting the queries from the server to each of the remotely programmable apparatuses;
  - f) transmitting the responses of the individuals to the queries from the remotely programmable apparatuses to the server;
  - g) creating a database of the individuals' behaviors and environments;
  - h) distinguishing a group of individuals having similar disease progressions;
  - i) using data mining techniques to categorize the group of individuals into at least two categories according to the individuals behavioral and environmental profiles;
  - j) determining at least a portion of the genotypes of the at least two categories of individuals;
  - k) using data mining techniques to find a gene difference between the at least two categories of individuals based at least in part upon a gene difference between the at least a portion of the respective genotypes; and
  - l) identifying the disease-influencing gene.
- R2  
end*

19. A method for identifying a disease-influencing substance, the  
method comprising the steps of:
- a) selecting individuals having a risk factor for a disease;
  - b) creating queries regarding the individuals' behaviors and environments;
  - c) storing the queries on a server;
  - d) providing each of the individuals with a remotely programmable apparatus having a user interface for communicating the queries and for receiving responses, and having communication means for communicating with the server through a communication network;
  - e) transmitting the queries from the server to each of the remotely programmable apparatuses;
  - f) transmitting the responses of the individuals to the queries from the remotely programmable apparatuses to the server;
  - g) creating a database of the individuals' behaviors and environments;
  - h) determining a gene sequence of a genotype for each of [the genotypes of] the individuals;
  - i) distinguishing a group of the individuals having similar gene sequences [genotypes];

- R  
Card*
- j) categorizing the group of individuals into at least two categories according to their disease progressions; and
  - k) using data mining techniques to find a disease-influencing substance from the behavioral and environmental profiles between the at least two classes of individuals.

- AH*
24. A database and data processing system for finding a disease-influencing gene among individuals having a risk factor for a disease, the database and data processing system comprising:
- a) a server for storing queries regarding the individuals' behavior and environment and for storing the individuals' responses to the queries;
  - b) at least one remotely programmable apparatus in communication with the server, wherein the remotely programmable apparatus comprises:
    - i) a user interface for communicating the queries to the individuals and for receiving the responses; and
    - ii) communication means for receiving the queries from the server and for transmitting the responses to the server;

- A  
X  
and*
- c) genotyping means in communication with the server for obtaining at least a portion of the genotype of the individual sufficient to group individuals having a similar genotype; and
  - d) data mining means in communication with the server, wherein the data mining means includes:
    - i) means for analyzing the responses in order to group the individuals having a similar behavioral and environmental profile, a similar disease progression, and a similar genotype;
    - ii) means for analyzing the responses in order to group the individuals having a similar disease progression;
    - iii) means for analyzing the responses in order to group the individuals having a similar genotype; and
    - iv) means for identifying the disease-influencing gene.
- 

- X*
27. A database and data processing system for use in finding a disease-influencing substance among individuals having a risk factor for a disease, the database and data processing system comprising:

- a) a server for storing queries regarding the individuals' behavior and environment and for storing the individuals' responses to the queries;
- b) at least one remotely programmable apparatus in communication with the server, wherein the remotely programmable apparatus comprises:
- i) a user interface for communicating the queries to the individuals and for receiving the responses; and
  - ii) communication means for receiving the queries from the server and for transmitting the responses to the server;
- c) genotyping means in communication with the server for obtaining at least a portion of the genotype of the individual; and
- d) data mining means in communication with the server, wherein the data mining means includes:
- i) means for analyzing the responses in order to group the individuals having a similar behavioral and environmental profile, a similar disease progression, and a similar at least a portion of the genotype;